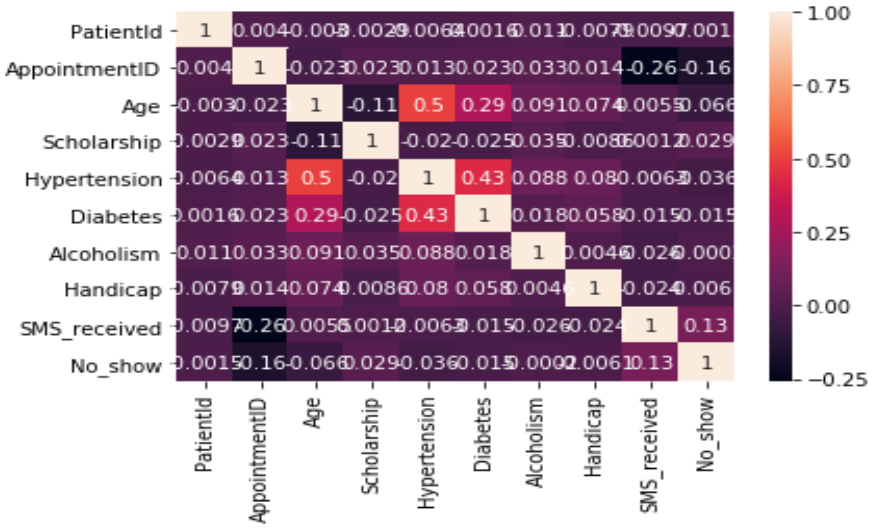
Project: No-show appointments

This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment. A number of characteristics about the patient are included in each row which will be analyzed by wrangling and exploring to get findings and results as shown.

After importing the needed libraries, gather, assess and cleaning data from some quality and tidiness issues, the main question was found about the main factors that affected on patient's showing up.

Exploring data can be detected by many way

From *heatmap*:



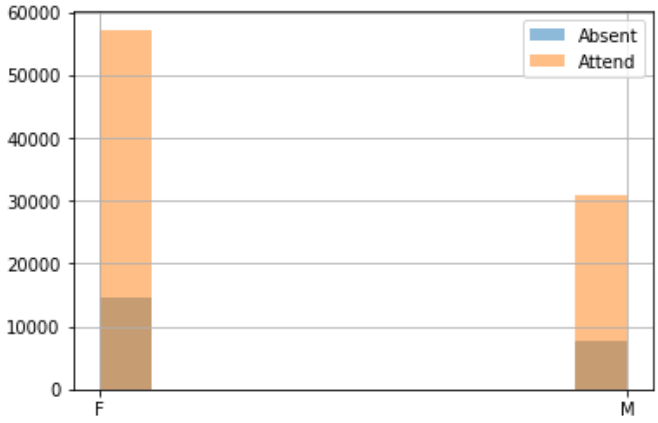
And the *described* statistics:

	PatientId	AppointmentID	Age	Scholarship	Hypertension	Diabetes	Alcoholism	Handicap	SMS_received	No_show
count	1.105270e+05	1.105270e+05	110527.000000	110527.000000	110527.000000	110527.000000	110527.000000	110527.000000	110527.000000	110527.000000
mean	1.474963e+14	5.675305e+06	38.273607	0.098266	0.197246	0.071865	0.030400	0.022248	0.321026	0.201933
std	2.560949e+14	7.129575e+04	22.105004	0.297675	0.397921	0.258265	0.171686	0.161543	0.466873	0.401444
min	3.921784e+04	5.030230e+06	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.172614e+12	5.640286e+06	20.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	3.173184e+13	5.680573e+06	37.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	9.439172e+13	5.725524e+06	55.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
max	9.999816e+14	5.790484e+06	115.000000	1.000000	1.000000	1.000000	1.000000	4.000000	1.000000	1.000000

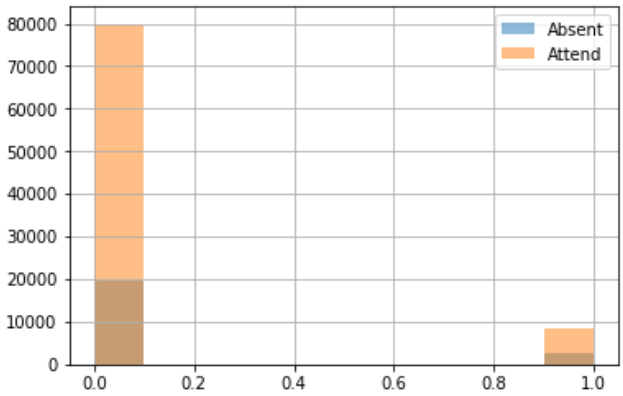
We can find that *No_show* is considered as dependent variable and there are many independent variables as shown in next relationships.

Showing up relationships:

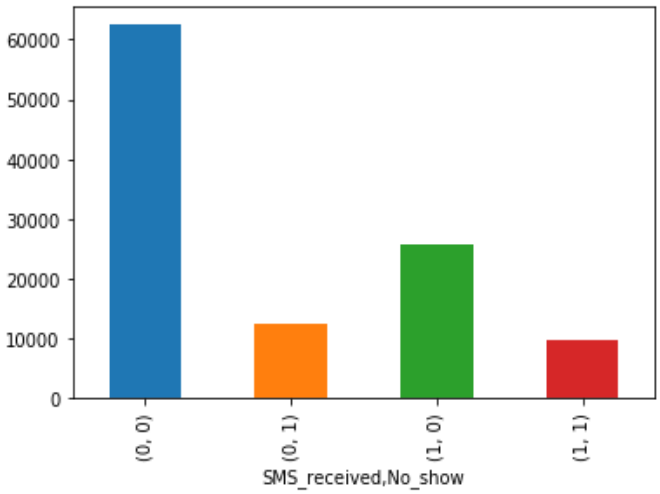
Gender: most females were attending to their appointments unlike males who had higher percentage to be absent according their assigned number compared with females.



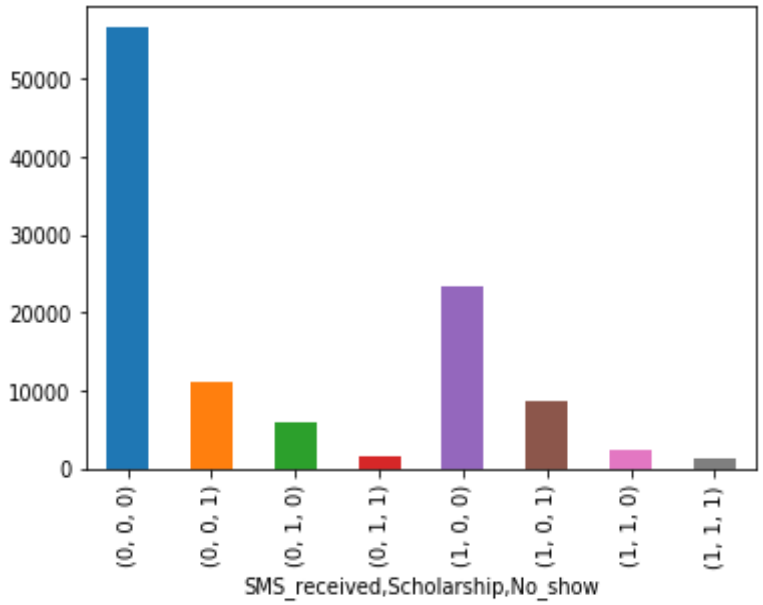
Scholarship: most of attended patients didn't enroll in scholarships.



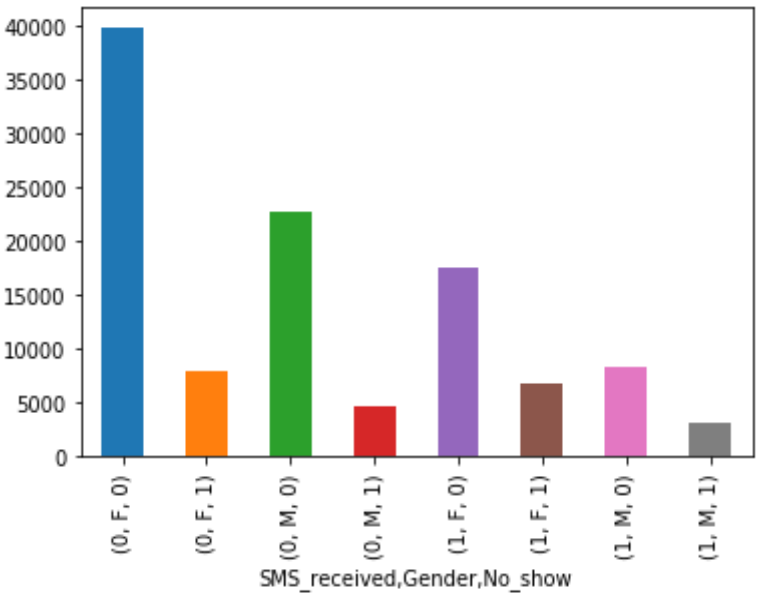
SMS received: the attended patients who didn't receive SMS were more than who already received SMS and for the absent patients the most of them didn't receive SMS



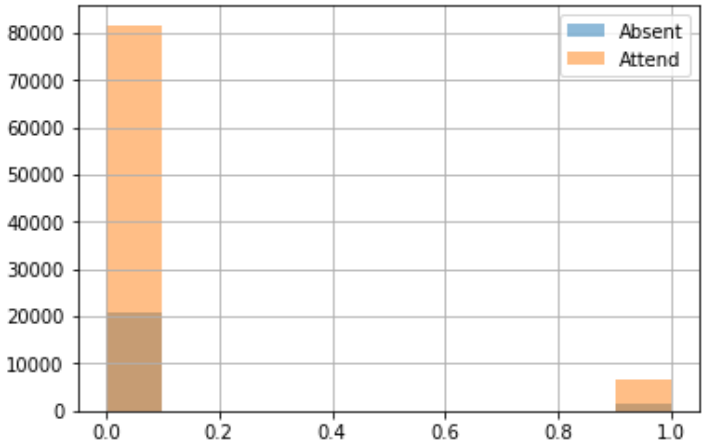
So, by grouping by both SMS receiving and enrolling in scholarship, we can find that the higher percentage of attendance was for who didn't receive SMS and didn't enroll in scholarships, after them who just received SMS



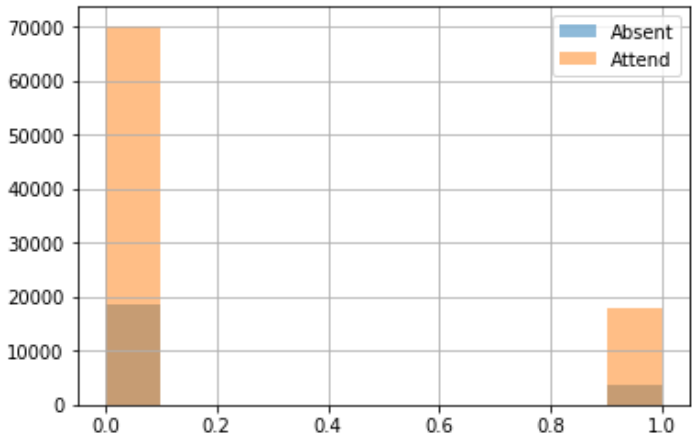
And by grouping by SMS receiving and gender to detect the showing up



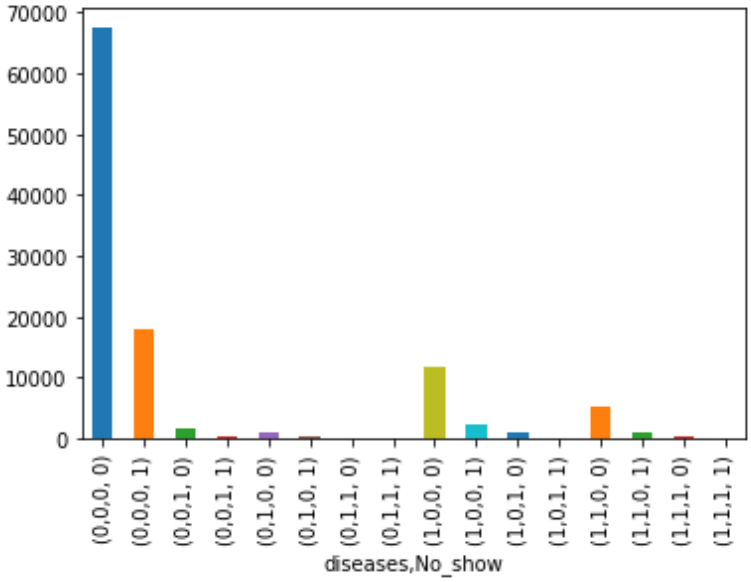
Diabetes: most of attended patients didn't have diabetes



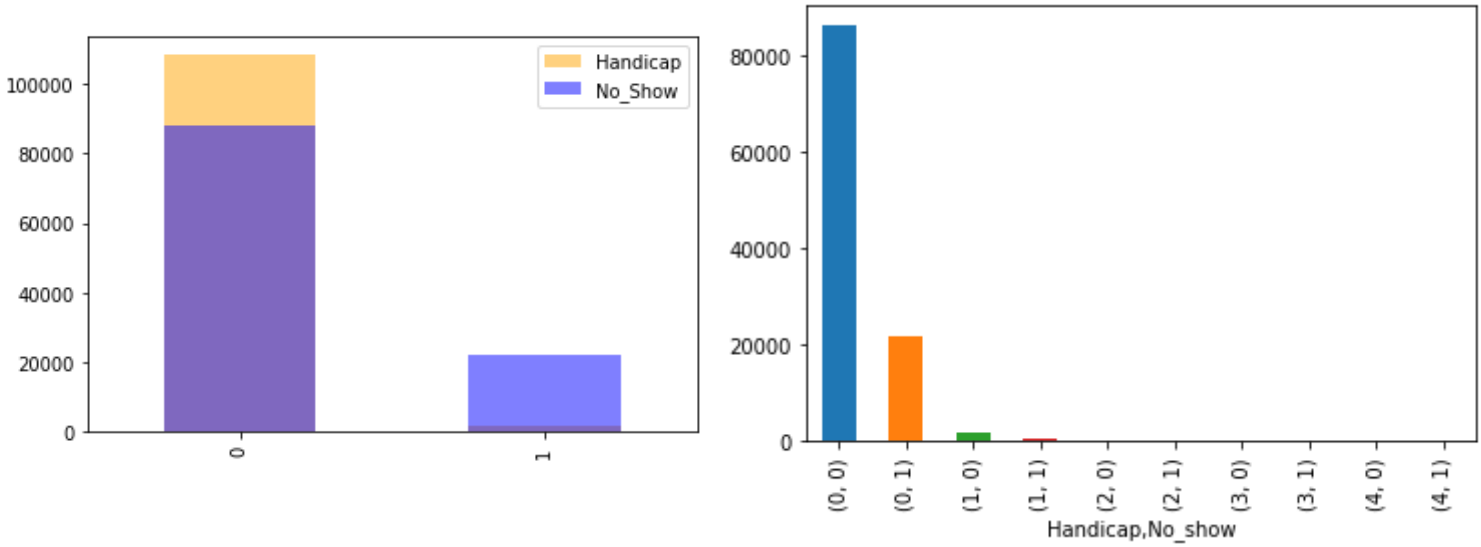
Hypertension: most of attended patients who didn't have hypertension



So, most of attended patients didn't have hypertension, diabetes and didn't alcoholism



Patients didn’t have any level of **handicaps** were more highly attend and a bite of absent patients were handicapped



After pervious analysis we can find that some of important factors to predict if the patients will show up to their appointments:

Gender as females appeared more committed than males, ***SMS receiving*** and ***Scholarships*** aren’t big deal as even the most percentage of attending patient were who didn’t received SMS or enroll in scholarship but they are still factors to make the showing up patients more as they linear proportional with patient attendance, about mention diseases like ***Hypertension***, ***Diabetes*** and ***Alcoholism*** they are linear proportional too with showing up as the most percentage of attendance was for who didn’t have any of those diseases and for ***handicapped*** patients by any levels were the most absent percentages unlike those who didn’t have handicap were more highly attend.